

VBench: Comprehensive Benchmark Suite for Video Generative Models

Year: 2023 | Citations: 866 | Authors: Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si

Abstract

Video generation has witnessed significant advancements, yet evaluating these models remains a challenge. A comprehensive evaluation benchmark for video generation is indispensable for two reasons: 1) Existing metrics do not fully align with human perceptions; 2) An ideal evaluation system should provide insights to inform future developments of video generation. To this end, we present VBench, a comprehensive benchmark suite that dissects “video generation quality” into specific, hierarchical, and disentangled dimensions, each with tailored prompts and evaluation methods. VBench has three appealing properties: 1) Comprehensive Dimensions: VBench comprises 16 dimensions in video generation (e.g., subject identity inconsistency, motion smoothness, temporal flickering, and spatial relationship, etc.). The evaluation metrics with fine-grained levels reveal individual models' strengths and weaknesses. 2) Human Alignment: We also provide a dataset of human preference annotations to validate our benchmarks' alignment with human perception, for each evaluation dimension respectively. 3) Valuable Insights: We look into current models' ability across various evaluation dimensions, and various content types. We also investigate the gaps between video and image generation models. We will open-source VBench, including all prompts, evaluation methods, generated videos, and human preference annotations, and also include more video generation models in VBench to drive forward the field of video generation.